AD_____

AWARD NUMBER: W81XWH-05-1-0391

TITLE: Maintaining Genome Stability: The Role of Helicases and Deaminases

PRINCIPAL INVESTIGATOR: Xiaojiang Chen

CONTRACTING ORGANIZATION: University of Southern California
Los Angeles, California 90089-1147

REPORT DATE: July 2006

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-07-2006	Annual	1 Jul 2005 – 30 Jun 2006

4. TITLE AND SUBTITLE

Maintaining Genome Stability: The Role of Helicases and Deaminases

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-05-1-0391

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Xiaojiang Chen

E-Mail: Xiaojiang.Chen@usc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Southern California
Los Angeles, California 90089-1147

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

No abstract provided.

15. SUBJECT TERMS
No subject terms provided.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	9	**19b. TELEPHONE NUMBER** *(include area code)*

The Brief Description of The Four Aims For The Grant:

(Aim 1). Study the *in vitro* functions of MCM proteins from archaea and yeast cells using the genetically engineered protein constructs. In this aim, we will also extend our prior success in the X-ray structural studies of an N-terminal fragment of an archaea MCM by attempting to crystallize MCM proteins from yeast.

(Aim 2). Examine *in vivo* effects of helicase function and in particular MCM roles in maintaining genome integrity in response to damage. This aim will use existing and newly generated mutants, which can be achieved through genetic screening and site-directed mutagenesis based on the 3-dimensional structure of MCM, to investigate how MCMs contribution to genome stability during chemical damage.

(Aim 3). Express, purify and crystallize the proteins of deaminases. We will focus on AID and APOBEC3G to obtain purified deaminase proteins for the *in vitro* biochemical, functional, and structural studies.

(Aim 4). Examine the functions and substrate specificity of AID and identify other factors required for the coupling of deamination with other processes of DNA synthesis and RNA transcription. The experiments will be carried out in a cell free assay system, using already purified DNA replication/repair and RNA transcription proteins in our labs.

Progress for Aim 1 (*in vitro* function of MCM):

The confirmation of double hexamer formation of the wt MCM: Even though biochemical and crystallographic evidence indicate that mtMCM assembles as a dHex (dHex), previous electron microscopy (EM) studies reported only the presence of single heptamers or single hexamers. Thus, there is controversy as to what are the oligomeric forms for the full length mtMCM. To further investigate this issue, we collaborated with Dr. Jose-Maria Carazo to study the EM structure of mtMCM. We observed dHex structure for the first time for the full-length (FL) mtMCM using 3D-EM reconstruction, in which two hexamers contact each other via the structurally well-defined N-terminal domains (Fig. 1A). This organization confirms our crystal structure of the N-mtMCM, which can fit into the EM density very well (Fig. 1A, 1C). The dHex has obvious side openings that resemble the side-channels of large T antigen (Fig. 1A).

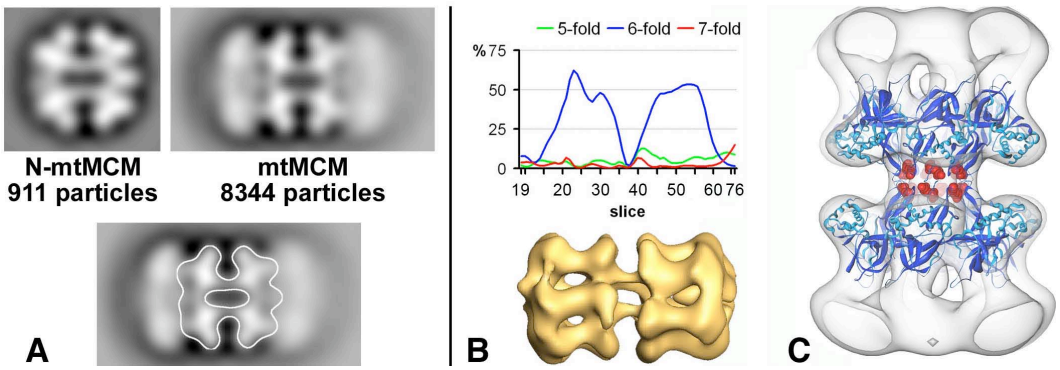

Fig. 1. 3D-EM reconstruction of mtMCM complex. (A) The EM density of N-mtMCM (top left) and full length (FL) mtMCM (top right). The N-mtMCM EM density can fit into that of the FL-mtMCM (bottom). (B) The symmetry analysis shows a hexameric symmetry on the horizontal axis, indicating a dHex structure. (C) The fitting of the X-ray model of the N-mtMCM into the FL-mtMCM.

The long, highly positively charged channel for DNA binding. Our recent crystal structure reveals that the dHex has a central channel of 118 Å in length from end to end (Fig. 2A). The channel has a diameter ranging from 47 Å at the widest point to 23 Å at the narrowest point, as measured from side chain to side chain (Fig. 2B). There are many positively charged residues located on the inner surface of the channel, generating a strongly positive channel; in contrast, the outside surface is mostly negative (Fig. 2A,2B). The charge distribution immediately suggests a DNA binding position.

DNA binding activity of the dHex. In addition to the positively charged channel of N-mtMCM, molecular modeling shows that the central channel of N-mtMCM is large enough to include ssDNA or dsDNA (Fig. 2C,2D). Indeed, the mtMCM protein was found to bind a 0.9 kb dsDNA fragment (Fig. 2E), supporting the hypothesis that mtMCM may be able to bind dsDNA in the central channel. This hypothesis is confirmed by our mutagenesis studies in which mutation of the residues on the β-hairpin inside the channel (see below).

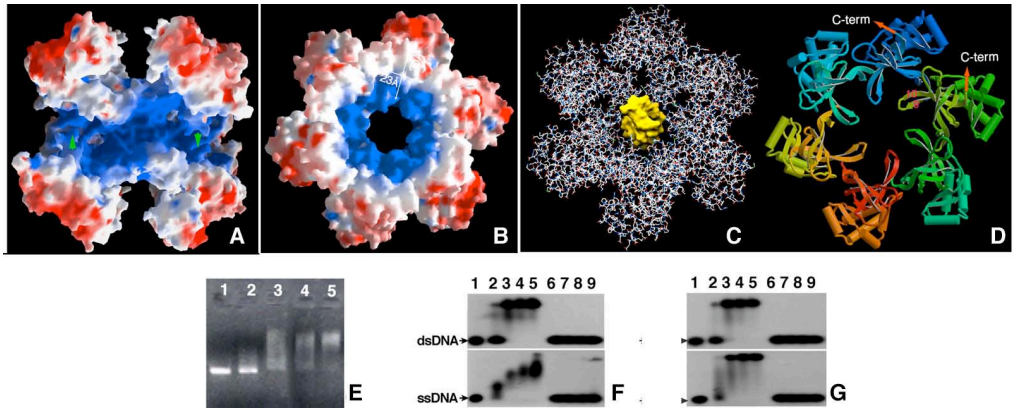

Fig. 2. The DNA binding studies of mtMCM. (A) The interior surface of the central channel of mtMCM, showing the highly positively charged surface (blue). Red is negatively charged surface, white is neutral. (B) Then end view of the charged channel. (C) The end view of the dHex of mtMCM, showing that dsDNA (yellow) fits the channel. (D) The b-hairpin structure inside the channel. We demonstrated that residues on the b-hairpin are critical for DNA binding and unwinding (see F, G). (E) DNA binding of mtMCM on native gel shift assay. 0.9 kb dsDNA fragment is shifted up with increasing amount of mtMCM added. (F, G). DNA binding of the β-hairpin mutants. The wt mtMCM binds DNA (lanes 1-5), but R226E (lanes 6-9 in panel-F) or K228E (lanes 6-9 in panel-G) mutants abolished dsDNA and ssDNA binding. These mutants also had no detectable helicase activity.

Biological activities of wt and single hexamer (sHex) mutants: We studied the enzymatic activities of the sHex mutants. We showed that the sHex mutants retained wild-type level of ATPase and DNA binding activities (data not shown), but had decreased helicase activity when compared with the wild type dHex protein (Fig.3). These biochemical properties of mtMCM are reminiscent of those of SV40 large T antigen, suggesting that the dHex form of mtMCM may be the active helicase for DNA unwinding during the bi-directional DNA replication.

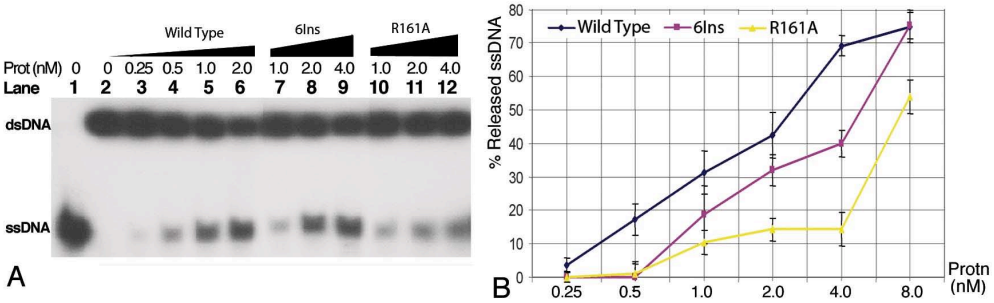

Fig. 3. Helicase activity of wt and sHex mutants of the mtMCM. (**A)** Gel analysis of helicase assay. Lane 1: boiled dsDNA substrate; lane 2: dsDNA substrate alone. The concentrations of wt and mutant proteins were indicated at the top. (**B)** Quantification of the helicase assay, graphed as percent ssDNA substrate released vs. protein concentrations that were calculated as dHex for wt and sHex for mutants. Error bars represent the average deviation of all assays quantified.

Progress for Aim 2

To examine *in vivo* effects of helicase function and in particular MCM roles in maintaining genome integrity in response to damage. This aim will use existing and newly generated mutants in MCM proteins to investigate how MCMs contribution to genome stability during chemical damage. We have initiated a screen for *mcm* mutants that are functional for DNA replication, but defective in response to damage. The screen design is as follows. We mutagenized *in vitro* plasmids containing wild type *mcm2*$^+$, *mcm4*$^+$, and *mcm7*$^+$. These were transformed into strains containing temperature sensitive mutations in *mcm2, mcm4* and *mcm7*, and selected for viability at 36°C, at which temperature the chromosomal MCM is inactivated. Cell viability depends upon the plasmid-borne MCM. This ensures that the plasmid copy is functional for replication. Then, we screened the surviving transformants for sensitivity to hydroxyurea.

Our first screen was conducted under conditions that are lethal for checkpoint mutants such as Δ*rad3*. We did not observe any hydroxyurea-sensitive MCM clones. However, following the old genetics adage "absence of evidence is not evidence of absence", we are re-calibrating our screening conditions to use a concentration of HU where we can see a phenotype in *Δcds1* checkpoint mutants. This strain is more resistant to HU, and our other experiments suggest that the Cds1 kinase may be involved directly in MCM function (see below), which makes it particularly relevant. We will also screen our transformants for sensitivity to other agents including MMS. Finally, we will screen existing *mcm* mutants using the same regimen.

To determine domains of MCMs responsible for interacting with other proteins, and investigate whether these interactions are mediated through single MCM subunits or the intact complex. As described in our preliminary data, we showed that the MCM proteins interact with the Rad51 (Rhp51) recombination protein. We have now shown that MCMs also interact with the Cds1 checkpoint kinase (human Chk2, budding yeast Rad53). These data are being submitted for publication. We hypothesize that MCMs may be a substrate of this kinase. In particular, we observe that an *mcm4ts* mutant blocked in hydroxyurea, and then released to restrictive temperature, is competent to complete replication and go on to divide. In contrast, this mutant shifted to restrictive temperature without pre-incubation in HU blocks late in S phase with severe damage. This suggests that HU has a protective effect on the *mcm4ts* allele. Further support of this comes from the same experiment with an *mcm4-degron* mutant. This allele causes rapid degradation of the Mcm4 protein at the restrictive temperature. As expected, the *mcm4-degron* mutant has the same phenotype regardless of pre-incubation with HU: a lethal arrest in S phase. We hypothesize that pre-activation of Cds1 by the HU treatment protects the Mcm4ts protein. We are now testing to see whether the temperature sensitive protein interacts with Cds1 and is a Cds1 substrate, and whether this influences interaction with other checkpoint proteins (e.g., Rad17).

Progress for Aim 3

In this aim, our plan is to express, purify and crystallize the proteins of deaminases, including AID, APOBEC3G, and other deaminaes that can be crystallized. We have tried extensively in E. coli expression of AID at first, but AID appeared to be very insoluble in E. coli expression, regardless of what deletion constructs and what expression and buffer conditions. Somehow, even in insect cell expression, we could not obtain large amount soluble AID protein that is sufficient for crystallization trials. We have subsequently tried to express and purify various APOBEC3 and APOBEC2 in *E. coli* and insect cells. Among the tested constructs, we found that one APOBEC2 deletion construct behaved well in solution, and we can purify sufficient amount for crystallization trials.

Using a Hampton screening system, we obtained some small crystals, which after optimization attempt, high diffraction quality crystals were obtained. The crystals diffracted to 2.5Å resolution in the space group of p212121. Selenium-methionine substituted proteins were made from minimal medium and Se-Met protein crystals were obtained. Se-MAD data were collected from ALS at Berkeley. The phases were solved using MAD as well as SADas using the program SOLVE. Currently, we have finished building the model and refined the structure. In the future, we plan to analyze the structure and carry out

the structure-based mutagenesis to understand the structure/function relationship of APOBEC protein family because the members of this family should have the conserved structure.

Progress for Aim4

Initiation of Human Hypermutation by the APOBEC Family of Nucleic Acid Deaminases: We have made considerable progress in dissecting the biochemical basis for the initiation of hypermutation by two members of the APOBEC family of nucleic acid cytidine deaminases, AID (activation-induced cytidine deaminase) and APOBEC3G. AID is required for somatic hypermutation and class-switch recombination of human immunoglobulin genes to generate high affinity antibodies. APOBEC3G is encapsulated along with the genomic RNA in the HIV-1 virion. APOBEC3G plays an important role in blocking infection by the HIV-1 (AIDS virus) in non-permissive cells. Both enzymes deaminate cytidine to form uracil (C → U) on ssDNA substrates. Almost all mutations are either inconsequential or seriously deleterious, but in accord with well-known "exception-that-proves-the-rule" principle, strictly regulated hypermuations driven by programmed nucleic acid deaminases are essential for human survival. Strict regulation of deamination is key, because unregulated deamination, caused by expression and subsequent action of deaminases in the wrong place at the wrong time are likely to lead to serious disease, e.g., AID-initiated B-cell lymphomas. Our primary goal is to study the biochemical and physical biochemical mechanisms for the APOBEC family of "mutators". The close collaboration of Xiaojiang Chen's laboratory with our laboratory, fostered by the Army Grant support, has enabled us to marry biochemical, physical biochemical analysis with structural analysis, and we have succeeded in obtaining a high resolution crystal structure of APOBEC2, which is present in heart and skeletal tissue.

Biochemical behavior of AID: We have shown that AID binds to randomly to ssDNA, irrespective of base composition, and catalyzes deamination of C favoring WRC (W = A/T, R = purine) hot spot motifs and disfavoring SYC (S = G/C, Y = pyrimidine) cold spot motifs. The enzyme scans processively along "naked" ssDNA and it appears to track processively on the non-transcribed strand of a moving transcription bubble. The deaminations occur in random clusters of between 4 and 10, usually containing at least one WRC triplet, interspersed with sparsely deaminated regions, i.e., regions with WRC motifs that are "ignored". The question is: what is the biochemical mechanism?

AID has an interesting N-terminal region, distal from its active site, which carries a large positive charge (+11). It seems likely that this region is important for binding to ssDNA. One and two site-directed mutations were made in this region, which reduced the charge to +9 and +7, respectively. The reduced charged mutants were less processive. Thus, as anticipated, there was a significant reduction in the number of deaminations/clone along with much fewer clusters of closely spaced deaminations. However, what we had not anticipated was the change in deamination specificity associated with the AID mutants R35E and R35E-R36D. The single mutant (R35E) strongly favored deamination in a non-WRC hot spot motif, where more than 70% of the clones had a mutation at this non-hot spot site, while the R35E-R36D double mutant retained the altered deamination specificity while additionally reducing mutations within a nearby hot spot motif.

We believe that the properties of the reduced charge AID mutants could turn out have fundamental importance, mechanistically and biologically. The biological significance is clear since there are human patients unable to synthesize high-affinity antibodies owing to replacement of Arg35. The mechanistic significance has two facets. First, the alteration in amino acid side-chain charge provides a straightforward opportunity to study how interactions with the negatively charged phosphate backbone influence the dynamics of AID movement. Second, and perhaps more importantly, the unexpected

change in deamination specificity could provide an opening to determine the role of AID phosphoryation mutational targeting, because the nearby Ser38 residue has been shown to be required for SHM and CSR *in vivo*. Therefore, the collaborative effort to obtain crystal structures for wild type and mutant AID should prove instrumental in revealing the mechanisms governing AID movement and deamination specificity, which in turn, would provide considerable insight into the mechanisms of immuno-diversity.

The effects of phosphorylation on AID activity

F. Alt and colleagues have established the importance of AID phosphorylation during SHM and CSR *in vivo*. We have compared phosphorylated wild type AID with dephosphorylated AID and AID mutants with mutations at predicted phosphorylation sites and have analyzed AID mutants with serine residues replaced by aspartic acid, which mimic phosphorylated AID (AID 'mimic mutants'). We have shown that AID purified from *baculovirus*-infected insect cells is phosphorylated at several serine residues, including the most important Ser38. Alt and coworkers showed that phosphorylation of Ser38 is required for B-cell AID to initiate SHM and CSR *in vivo* and for activity on transcribed dsDNA, but not ssDNA *in vitro*. Our data agree that Ser38 phosphorylation is needed for optimal AID activity, however, in contrast, we showed that both dephosphorylated AID and Ser38 to Ala mutant have a decreased specific activity on both ssDNA and transcribed dsDNA. The effect of phosphorylation appears principally to be electrostatic because replacement of Ser38 with Asp, "phosphorylation mimic mutation", yields fully active AID.

In contrast to some of Alt's data, we observed avid AID-catalyzed deamination during active transcription in a T7 RNA polymerase system, which did not require the presence of RPA (human single-stranded binding protein), whereas Alt's data showed that RPA is needed to deaminate linear dsDNA using human AID. The reason for this discrepancy isn't clear. Two possible explanations come to mind. Perhaps native human B cell AID behaves differently from recombinant human AID grown in *baculovirus*-infected insect cells. However, AID purified from insect cells exhibits bona fide SHM hallmarks – selective deamination at WRC hot spots, deamination avoidance at SYC cold spots, random clustered deaminations, transcription-dependent deamination and a proper phosphorylation pattern. We believe that a possible, perhaps likely, explanation is that AID from insect cells may be considerably more active than AID purified from B cells (based on comparative assays performed in our lab). The more active AID appears to track along the non-transcribed strand of a moving transcription bubble. A less active deaminase might require RPA to stabilize the ssDNA during transcription. Whether or not RPA has a role to play in transcription-dependent AID-catalyzed deamination, phosphorylation of AID at several residues, particularly at Ser38, is clearly very important.

Biochemical behavior of APOBEC3G

Like AID, APOBEC3G scans ssDNA processively. However, in stark contrast to AID, deamination is directional not random, and exhibits a strong 3' → 5' preference. Directional deamination is remarkable in the sense that there is no hydrolysis of ATP, and, if fact, the assay contains no nucleotide cofactors. We have proposed that similar to AID, APOBEC3G translocates in both directions, but can only catalyze deamination of C when translocating 3' to 5'. It is known that the main target motif for APOBEC3G is CCC. We have observed that it is the bases immediately adjacent to the 3'-side of CCC that determine the probability of APOBEC3G-catalyzed deamination. For example, APOBEC3G has a 20-fold higher specific activity for deamination of 5'CCCAAA3' compared to 5'CCCTTT3'. Thus, APOBEC3G appears to "see" the target CCC motif from the 3'-side. We further showed that APOBEC3G can both slide and jump along the ssDNA while deaminating a *single* ssDNA substrate. We have obtained preliminary data suggesting that AID can also perform processive sliding and jumping

motion on ssDNA – a process referred to as "facilitated diffusion", but while APOBEC catalyzes directional deamination, the action of AID is random.

The biological consequence of directional deamination is that it offers a plausible biochemical explanation for the well-established property of HIV-1, which has a much higher G → A mutational density toward the 3'-end of its RNA genome caused by having more deamination toward the 5'-end of cDNA by APOBEC3G.

Our recent paper in Nature Structural and Molecular Biology, Chelico et al., Publication 1, was cited as a Research Highlight in Nature, 440, 2006, p. 1093, under the heading Molecular Biology, "Slide and jump". The paper was also featured in a Nature Structural and Molecular Biology News and Views, 440, 380-381 (2006), RS Harris & H Matsuo, entitled "Dancin' deaminase".

Significance

The humoral response is able to recognize and attack an "unlimited" number of possible infectious agents by radically diversifying the inherited immunoglobulin (Ig) gene. First, the Ig gene is constructed from a variety of inherited gene segments by a Rag-mediated VDJ recombination event which joins together a variable (V), diversity (D) and joining (J) segments to make a functional Ig gene capable of expressing a low affinity antibody. In order to completely clear an infection, B cells undergo affinity maturation to produce antibodies with stronger affinity for antigen and with appropriate effector function. Affinity maturation involves two diversification events, somatic hypermutation and class switch recombination (SHM and CSR). Our principal long-term objective is to focus on SHM and to reconstitute the hypermutational process in an *in vitro* model system. Our initial objective is to explore the action of activation-induced cytidine deaminase (AID), an enzyme required for both SHM and CSR. We have recently shown that AID-catalyzed deamination of cytosine residues on ssDNA simulates three hallmark properties of SHM – mutational hot and cold spot sequences, broad clonal heterogeneity and transcriptional dependence. Active transcription is also a requirement for SHM, and our current data suggest that AID is able to target C → T mutations preferentially on the non-transcribed strand of a moving transcription bubble. An immediate priority of ours is to develop an *in vitro* human transcription system to investigate how AID is targeted to actively transcribed V- genes but not C genes. A biochemical understanding of AID, whose expression is confined to B cells, and the other APOBEC homologs, whose expression is diverse, should provide important new information on the potential risk factors, which we have referred to as "at risk nucleic acid sequences", which when acted upon inappropriately by APOBEC nucleic acid C deaminases can have serious deleterious consequences resulting from APOBEC-induced genomic instability, leading to cancer and perhaps even to neurodegenerative disease. Progress in understanding the molecular interactions governing SHM, which along with CSR is prerequisite to generating high affinity antibodies, would benefit substantially by the availability of an *in vitro* model system to study AID's role. We intend to provide this invaluable biochemical component.

Future Plans

Although we have progressed towards understanding the biochemical mechanisms responsible for initiating mutations by AID during SHM and by APOBEC3G during HIV-1 inactivation, an in-depth analysis will clearly require high-resolution structural analysis of wild type and mutant APOBEC family enzymes. During the next grant period, we will work in close collaboration with Xiaojiang Chen to obtain crystals suitable for X-ray analysis. We have successfully crystallized APOBEC2. APOBEC2 is found in cardiac and skeletal muscle tissue, but its biological role is currently unknown, even though it contains many of the conserved motifs needed for AID activity, including phosphorylation motifs. AID had "no" obvious activity until we succeeded in eliminating a tightly bound RNA inhibitor. Similarly, an RNA inhibitor is also associated with APOBEC3G, which must be removed for the enzyme to

deaminate ssDNA. Therefore, it is no surprise that APOBEC2 has an avidly bound RNA, but its removal has not revealed an activity for APOBEC2. Next, we will explore the possibility that metal ion cofactors or perhaps protein activators may be needed. In parallel with the study of APOBEC2, we are attempting to crystallize other APOBEC family members, APOBEC3G and AID, which we've analyzed biochemically, and APOBEC3F, about which little is known.

Publications

1. Chelico, L., Pham, P., Calabrese, P., and Goodman, M. F. Jumping and Sliding, APOBEC3G DNA Deaminase Acts Processively 3' to 5' on Single-Stranded DNA. Nat. Struct. Mol. Biol. 13, 392-400 (2006).
2. Bransteitter, R., Sneeden, J., Allen, S., Pham, P., Goodman, M. F. First AID (Activation-induced Cytidine Deaminase) is Needed to Produce High-affinity Isotype-switched Antibodies. J. Biol. Chem. 281, (2006, in press).